Edge Computing with Artificial Intelligence: A Machine Learning Perspective

Abstract

Recent years have witnessed the widespread popularity of Internet of things (IoT). By providing sufficient data for model training and inference, IoT has promoted the development of artificial intelligence (AI) to a great extent. Under this background and trend, the traditional cloud computing model may nevertheless encounter many problems in independently tackling the massive data generated by IoT and meeting corresponding practical needs. In response, a new computing model called edge computing (EC) has drawn extensive attention from both industry and academia. With the continuous deepening of the research on EC, however, scholars have found that traditional (non-AI) methods have their limitations in enhancing the performance of EC. Seeing the successful application of AI in various fields, EC researchers start to set their sights on AI, especially from a perspective of machine learning, a branch of AI that has gained increased popularity in the past decades. In this article, we first explain the formal definition of EC and the reasons why EC has become a favorable computing model. Then, we discuss the problems of interest in EC. We summarize the traditional solutions and hightlight their limitations. By explaining the research results of using AI to optimize EC and applying AI to other fields under the EC architecture, this article can serve as a guide to explore new research ideas in these two aspects while enjoying the mutually beneficial relationship between AI and EC.